FIG.1

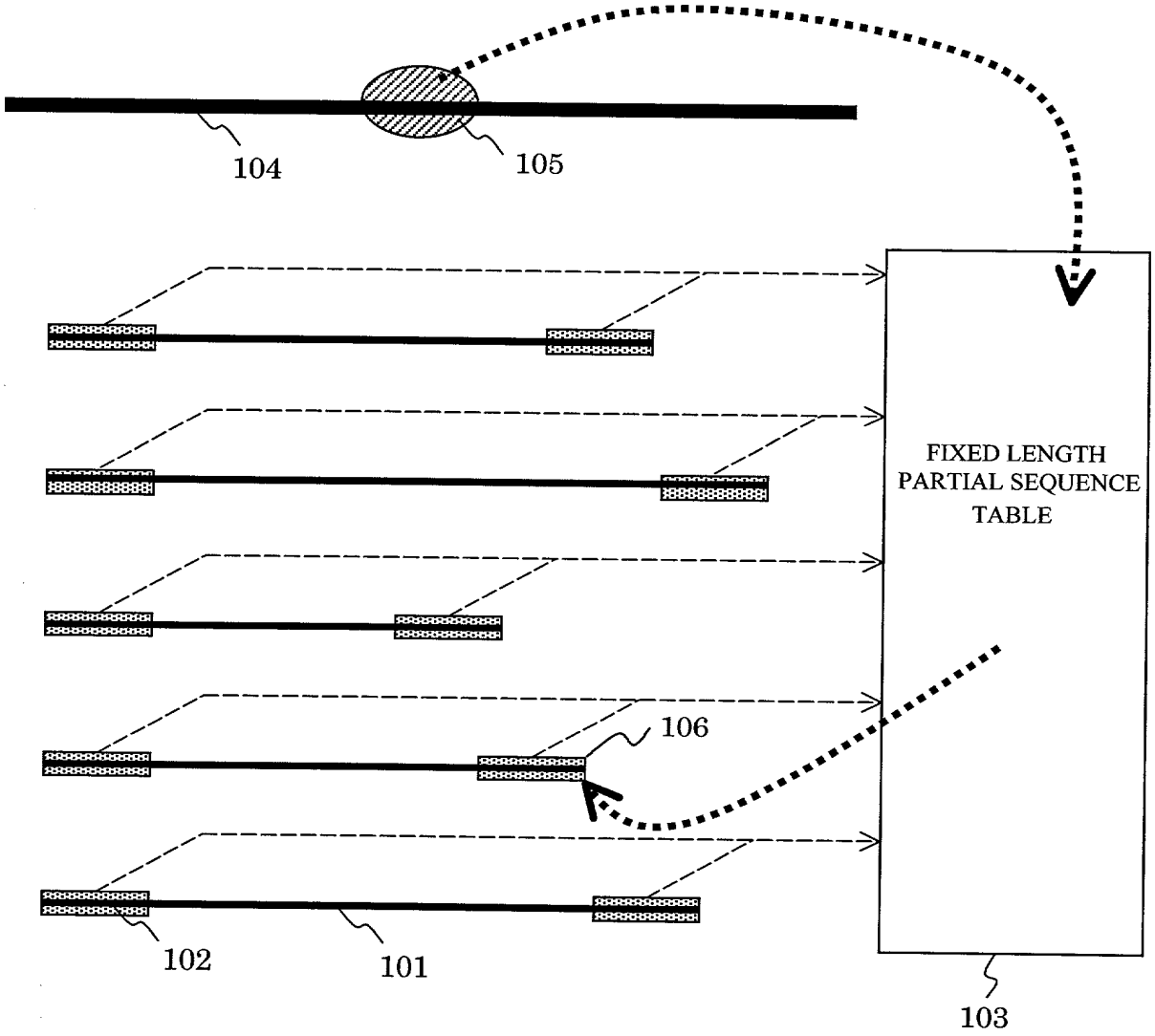


FIG.2

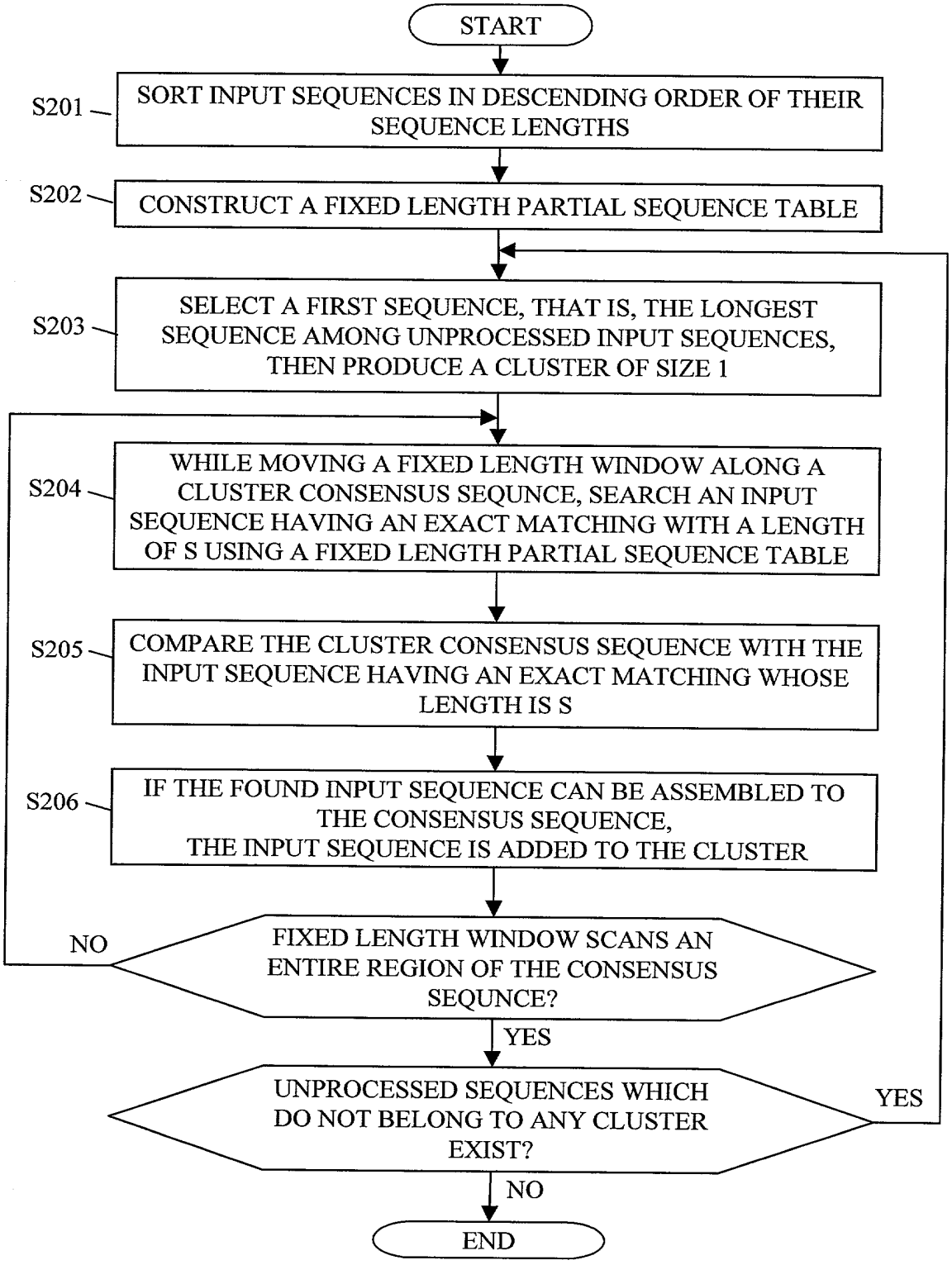

START
S201
SORT INPUT SEQUENCES IN DESCENDING ORDER OF THEIR SEQUENCE LENGTHS
S202
CONSTRUCT A FIXED LENGTH PARTIAL SEQUENCE TABLE
S203
SELECT A FIRST SEQUENCE, THAT IS, THE LONGEST SEQUENCE AMONG UNPROCESSED INPUT SEQUENCES, THEN PRODUCE A CLUSTER OF SIZE 1
S204
WHILE MOVING A FIXED LENGTH WINDOW ALONG A CLUSTER CONSENSUS SEQUNCE, SEARCH AN INPUT SEQUENCE HAVING AN EXACT MATCHING WITH A LENGTH OF S USING A FIXED LENGTH PARTIAL SEQUENCE TABLE
S205
COMPARE THE CLUSTER CONSENSUS SEQUENCE WITH THE INPUT SEQUENCE HAVING AN EXACT MATCHING WHOSE LENGTH IS S
S206
IF THE FOUND INPUT SEQUENCE CAN BE ASSEMBLED TO THE CONSENSUS SEQUENCE, THE INPUT SEQUENCE IS ADDED TO THE CLUSTER
NO
FIXED LENGTH WINDOW SCANS AN ENTIRE REGION OF THE CONSENSUS SEQUNCE?
YES
UNPROCESSED SEQUENCES WHICH DO NOT BELONG TO ANY CLUSTER EXIST?
YES
NO
END

FIG.3

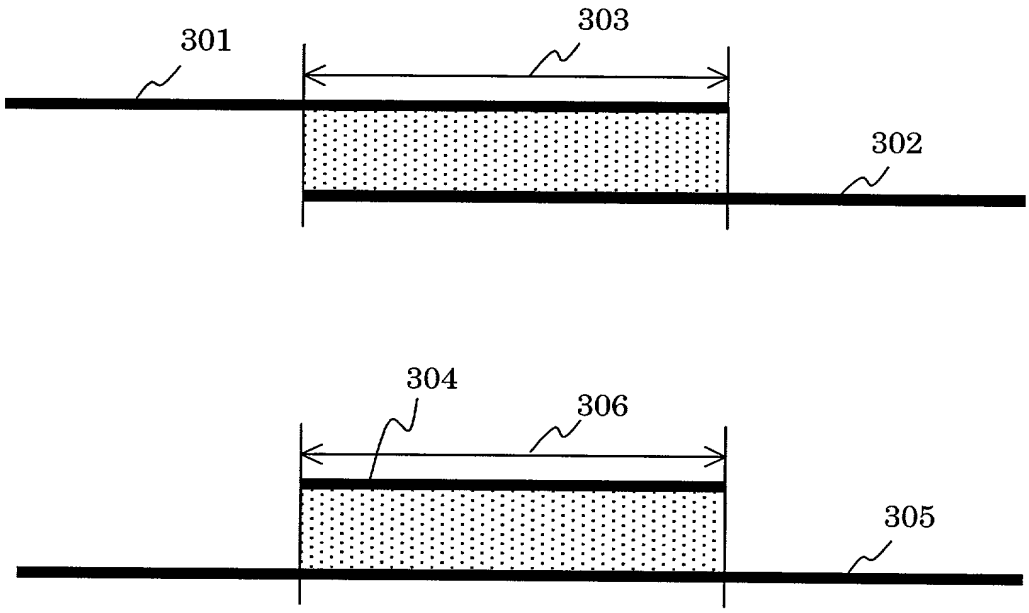
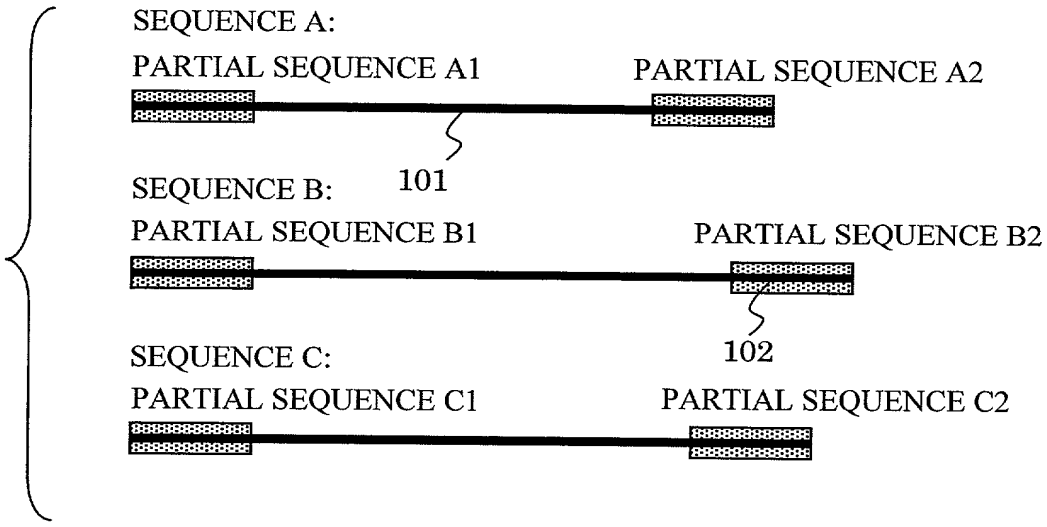
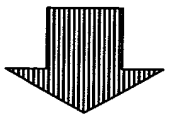


FIG.4





103

PARTIAL SEQUENCE	INPUT SEQUENCE	POSITION
A1	A	0
A2	A	400
B1	B	0
B2	B	500
C1	C	0
C2	C	450

401

402

FIG.5

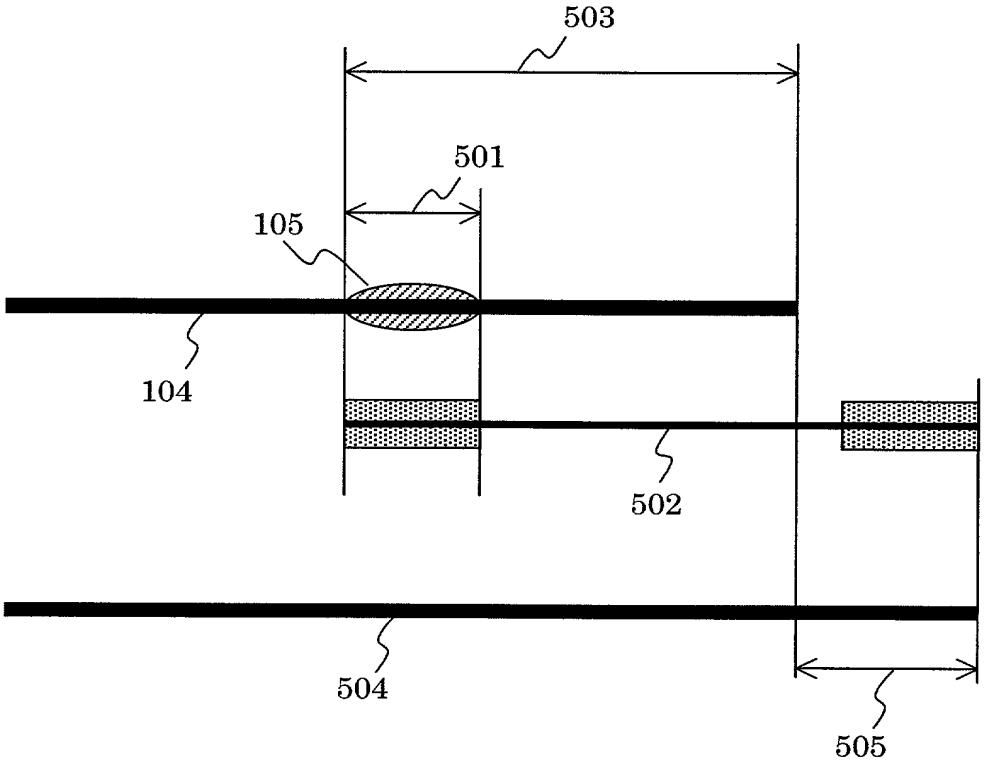


FIG.6

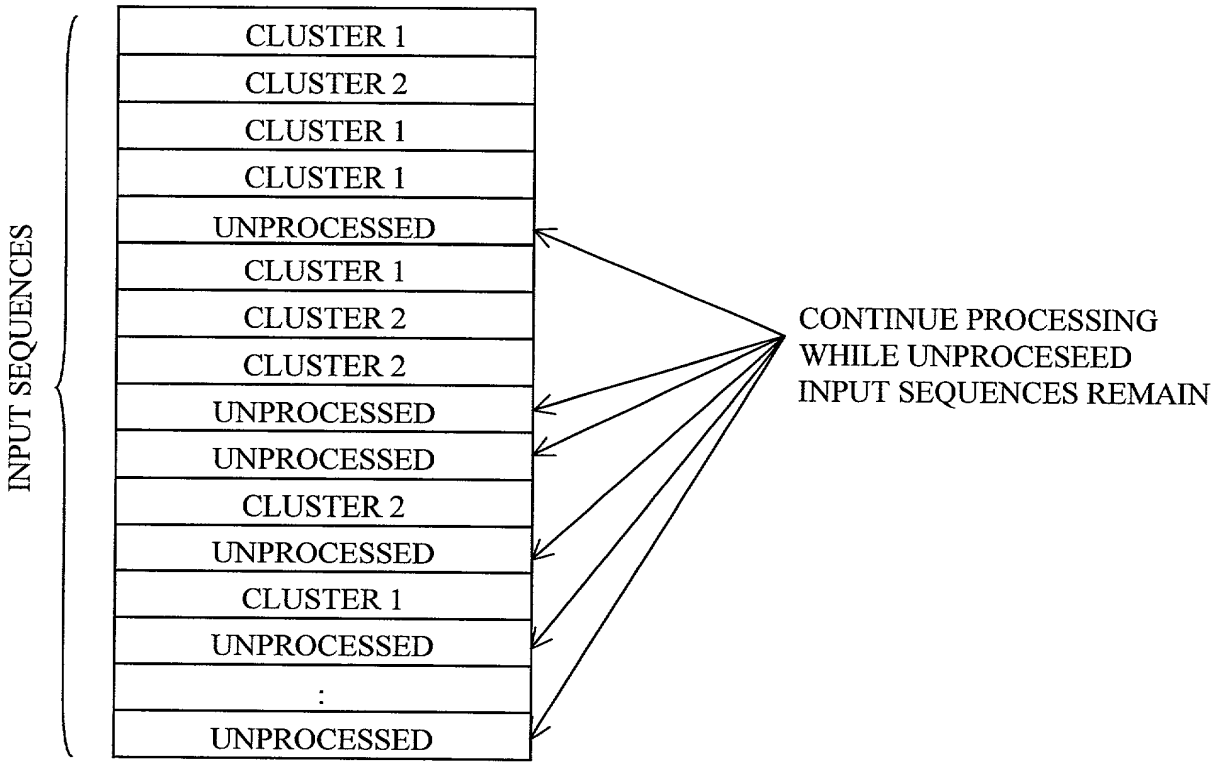
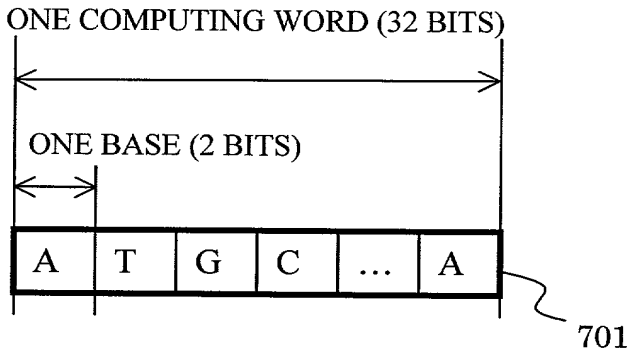

INPUT SEQUENCES
CLUSTER 1
CLUSTER 2
CLUSTER 1
CLUSTER 1
UNPROCESSED
CLUSTER 1
CLUSTER 2
CLUSTER 2
UNPROCESSED
UNPROCESSED
CLUSTER 2
UNPROCESSED
CLUSTER 1
UNPROCESSED
:
UNPROCESSED
CONTINUE PROCESSING
WHILE UNPROCESEED
INPUT SEQUENCES REMAIN

FIG.7

IF ONE COMPUTING WORD IS USED



IF TWO COMPUTING WORDS ARE USED

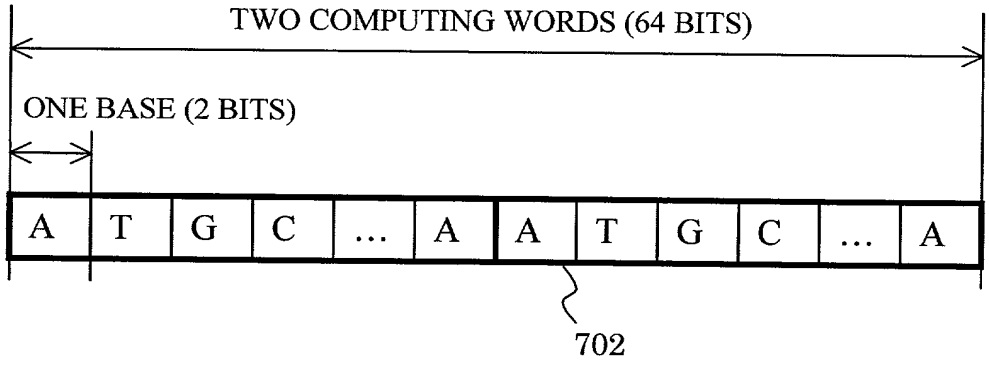


FIG.8

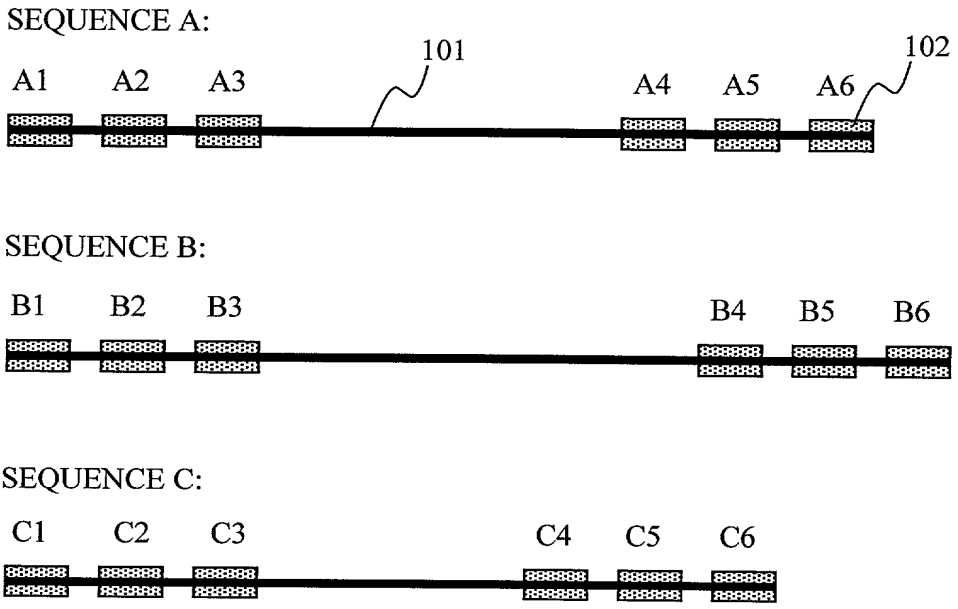

SEQUENCE A:
A1
A2
A3
101
A4
A5
A6
102
SEQUENCE B:
B1
B2
B3
B4
B5
B6
SEQUENCE C:
C1
C2
C3
C4
C5
C6

FIG.9

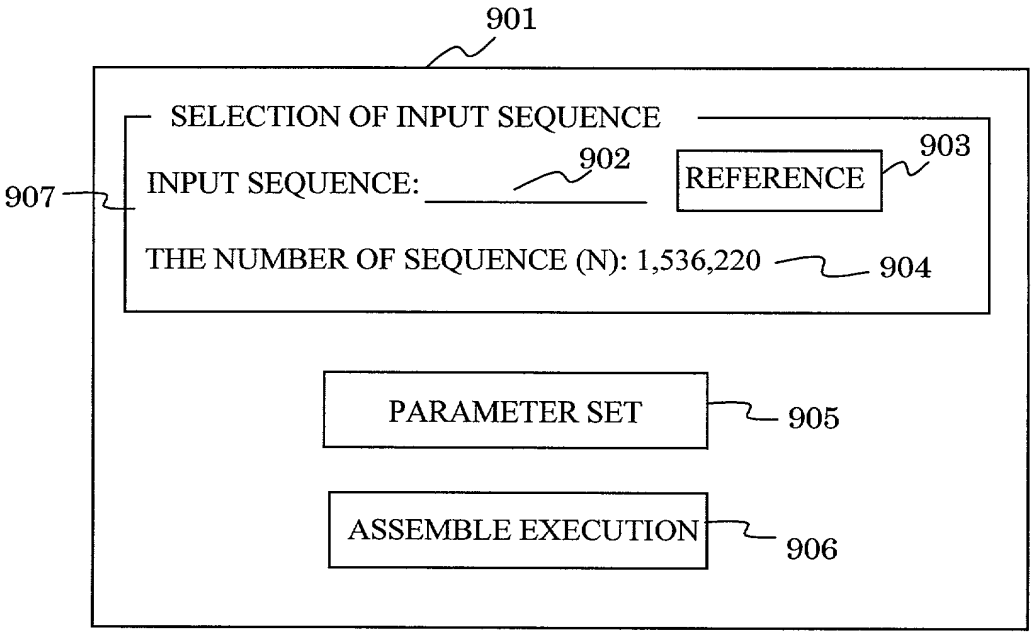

901
SELECTION OF INPUT SEQUENCE
907
INPUT SEQUENCE:
902
REFERENCE
903
THE NUMBER OF SEQUENCE (N): 1,536,220
904
PARAMETER SET
905
ASSEMBLE EXECUTION
906

FIG.10

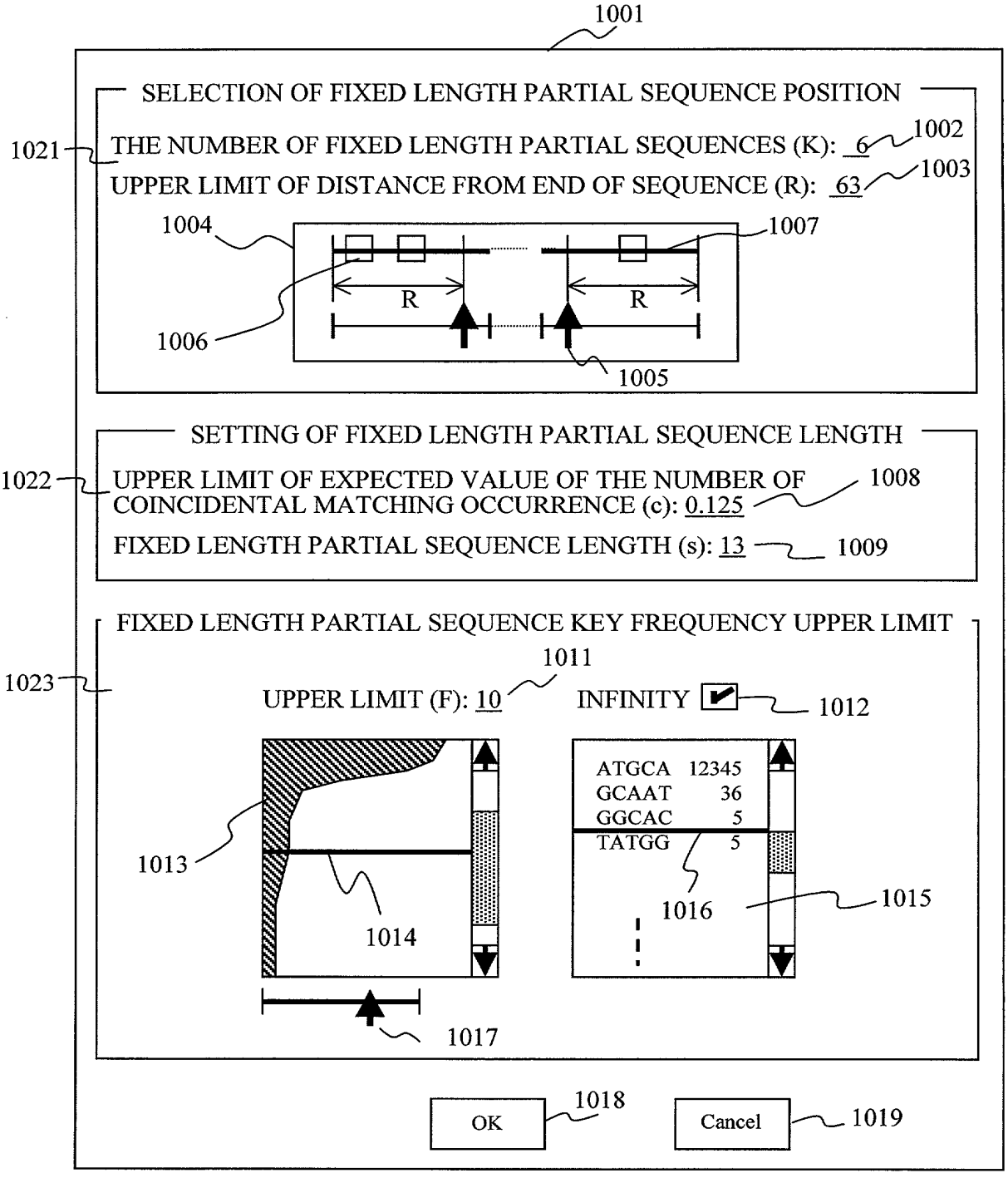

1001
SELECTION OF FIXED LENGTH PARTIAL SEQUENCE POSITION
1021
THE NUMBER OF FIXED LENGTH PARTIAL SEQUENCES (K): 6
1002
UPPER LIMIT OF DISTANCE FROM END OF SEQUENCE (R): 63
1003
1004
1007
R
R
1006
1005
SETTING OF FIXED LENGTH PARTIAL SEQUENCE LENGTH
1022
UPPER LIMIT OF EXPECTED VALUE OF THE NUMBER OF COINCIDENTAL MATCHING OCCURRENCE (c): 0.125
1008
FIXED LENGTH PARTIAL SEQUENCE LENGTH (s): 13
1009
FIXED LENGTH PARTIAL SEQUENCE KEY FREQUENCY UPPER LIMIT
1023
1011
UPPER LIMIT (F): 10
INFINITY
1012
1013
1014
ATGCA 12345
GCAAT 36
GGCAC 5
TATGG 5
1016
1015
1017
OK
1018
Cancel
1019

FIG.11

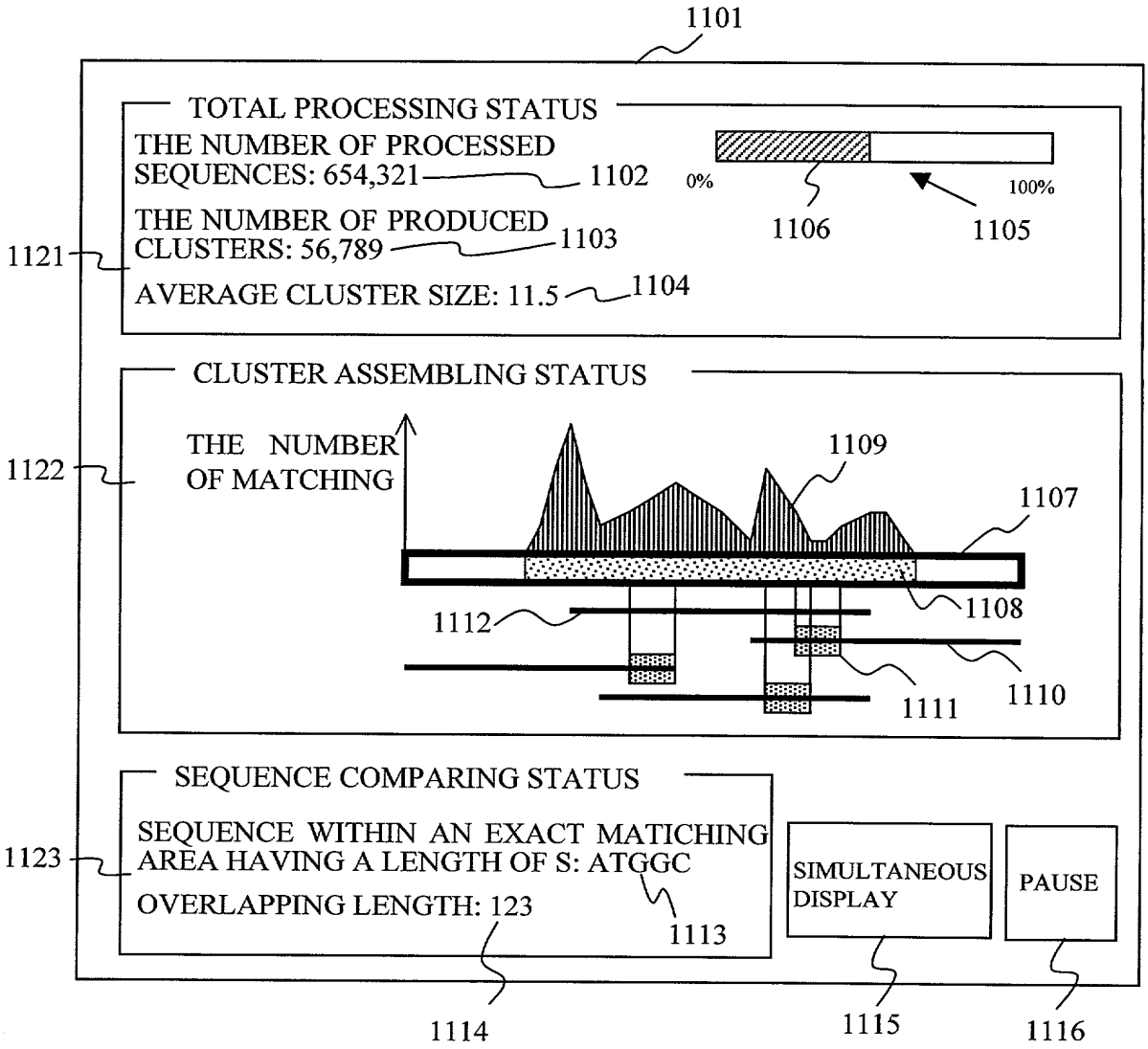

1101
TOTAL PROCESSING STATUS
THE NUMBER OF PROCESSED SEQUENCES: 654,321
1102
0%
100%
1106
1105
THE NUMBER OF PRODUCED CLUSTERS: 56,789
1103
1121
AVERAGE CLUSTER SIZE: 11.5
1104
CLUSTER ASSEMBLING STATUS
1122
THE NUMBER OF MATCHING
1109
1107
1108
1112
1111
1110
SEQUENCE COMPARING STATUS
1123
SEQUENCE WITHIN AN EXACT MATICHING AREA HAVING A LENGTH OF S: ATGGC
1113
OVERLAPPING LENGTH: 123
SIMULTANEOUS DISPLAY
PAUSE
1114
1115
1116

FIG.12

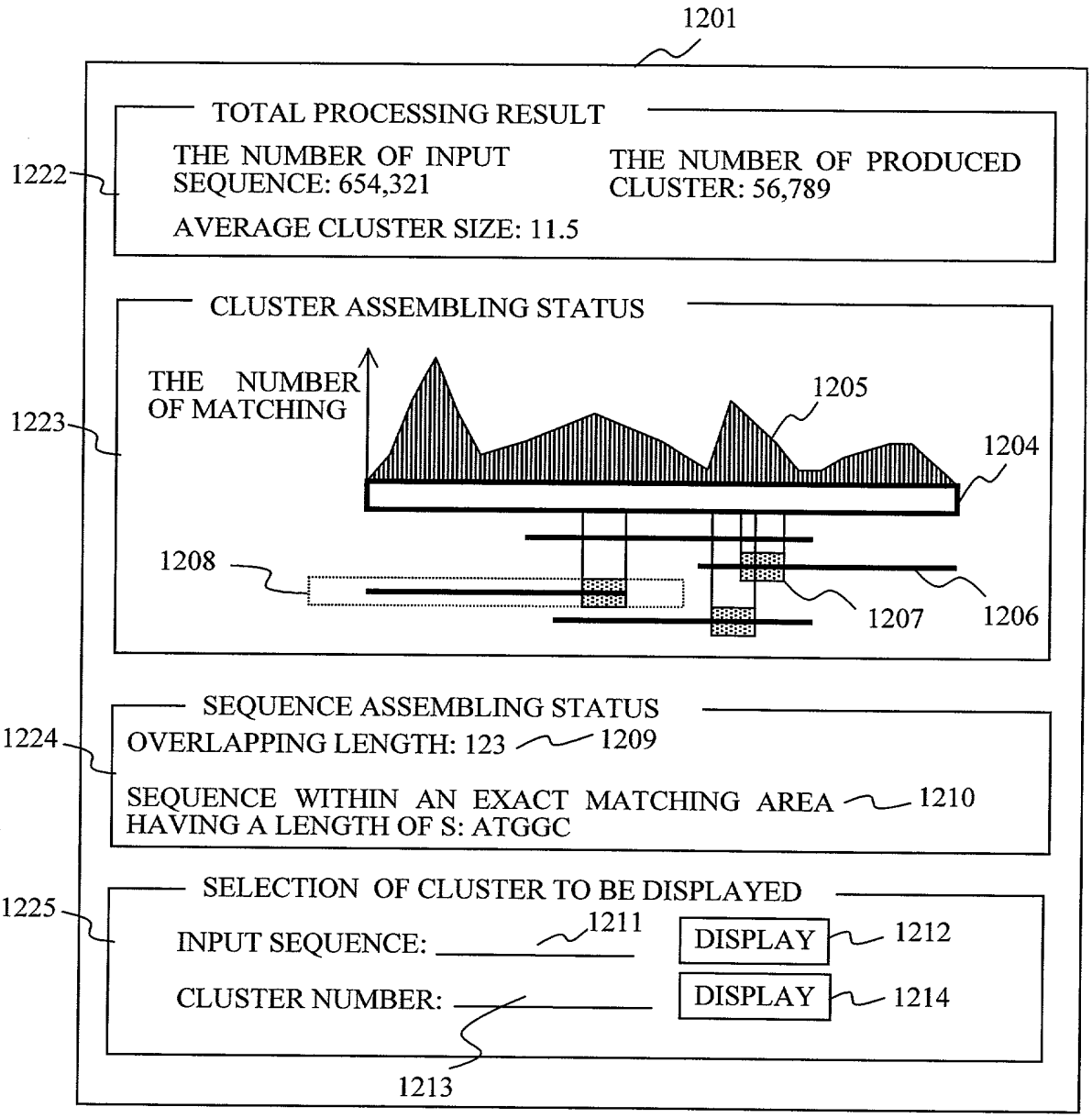

1201
TOTAL PROCESSING RESULT
THE NUMBER OF INPUT SEQUENCE: 654,321
THE NUMBER OF PRODUCED CLUSTER: 56,789
AVERAGE CLUSTER SIZE: 11.5
1222
CLUSTER ASSEMBLING STATUS
THE NUMBER OF MATCHING
1205
1204
1208
1207
1206
1223
SEQUENCE ASSEMBLING STATUS
OVERLAPPING LENGTH: 123
1209
SEQUENCE WITHIN AN EXACT MATCHING AREA HAVING A LENGTH OF S: ATGGC
1210
1224
SELECTION OF CLUSTER TO BE DISPLAYED
INPUT SEQUENCE:
1211
DISPLAY
1212
CLUSTER NUMBER:
DISPLAY
1214
1225
1213

FIG.13

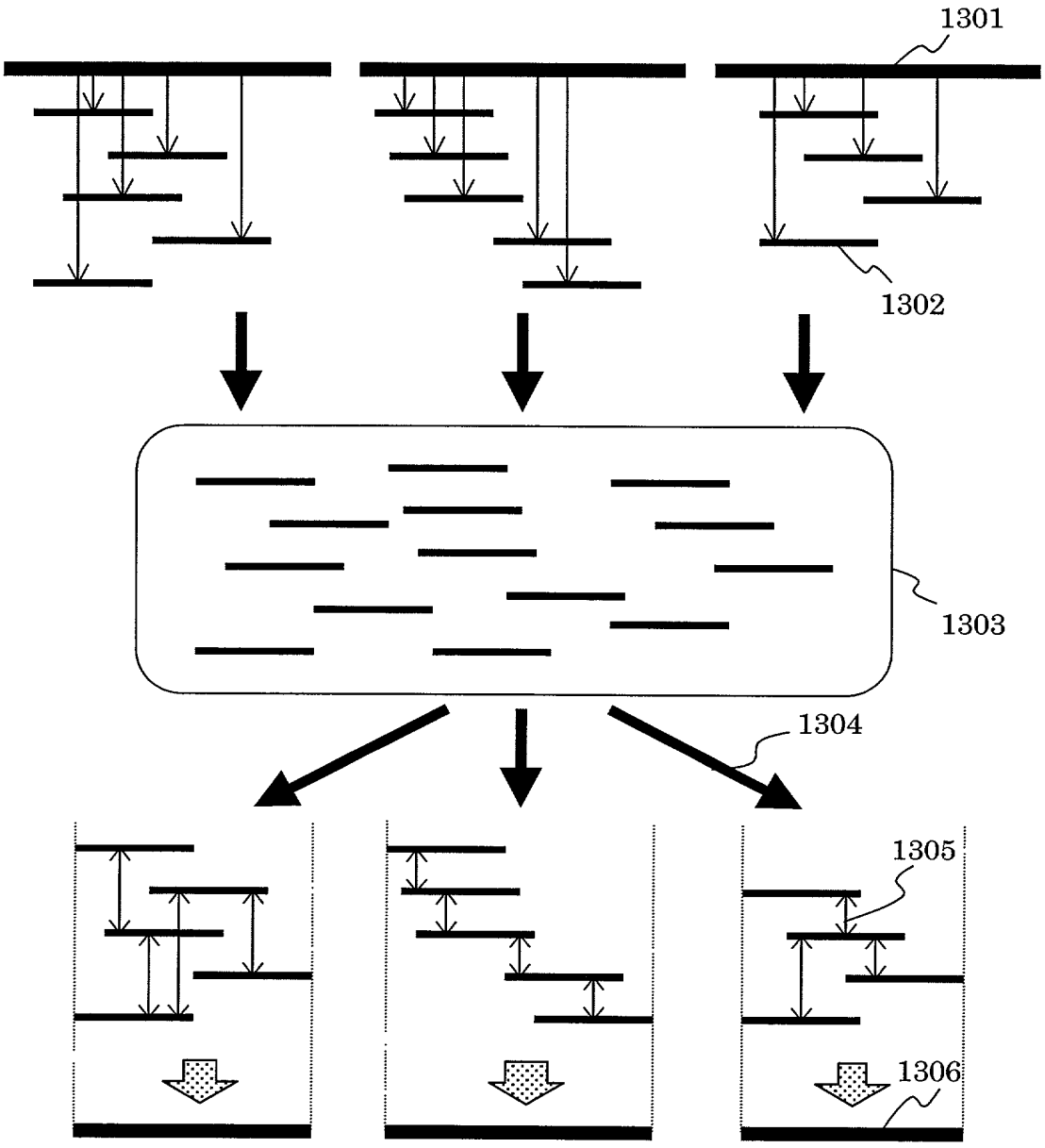


FIG.14

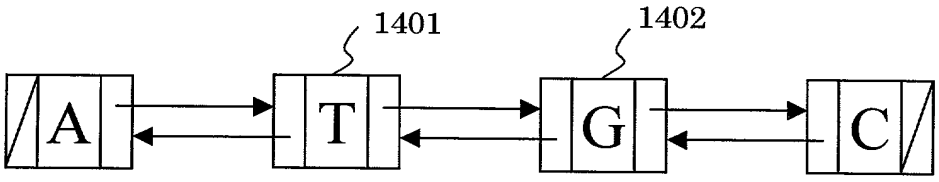


FIG.15

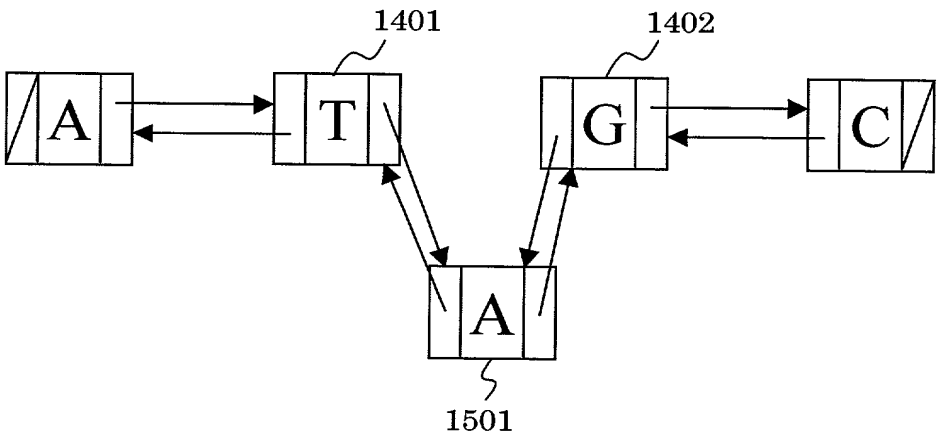


FIG.16

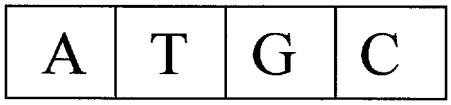


FIG.17

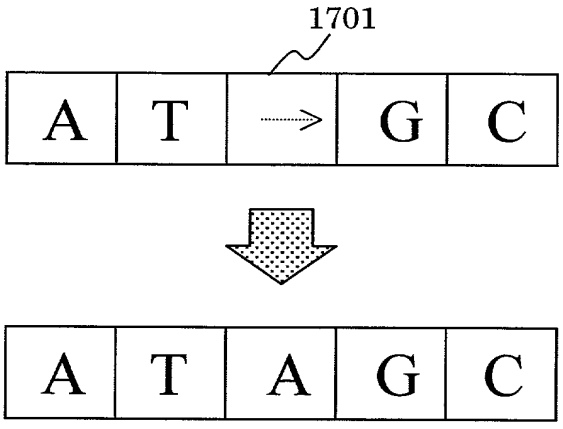
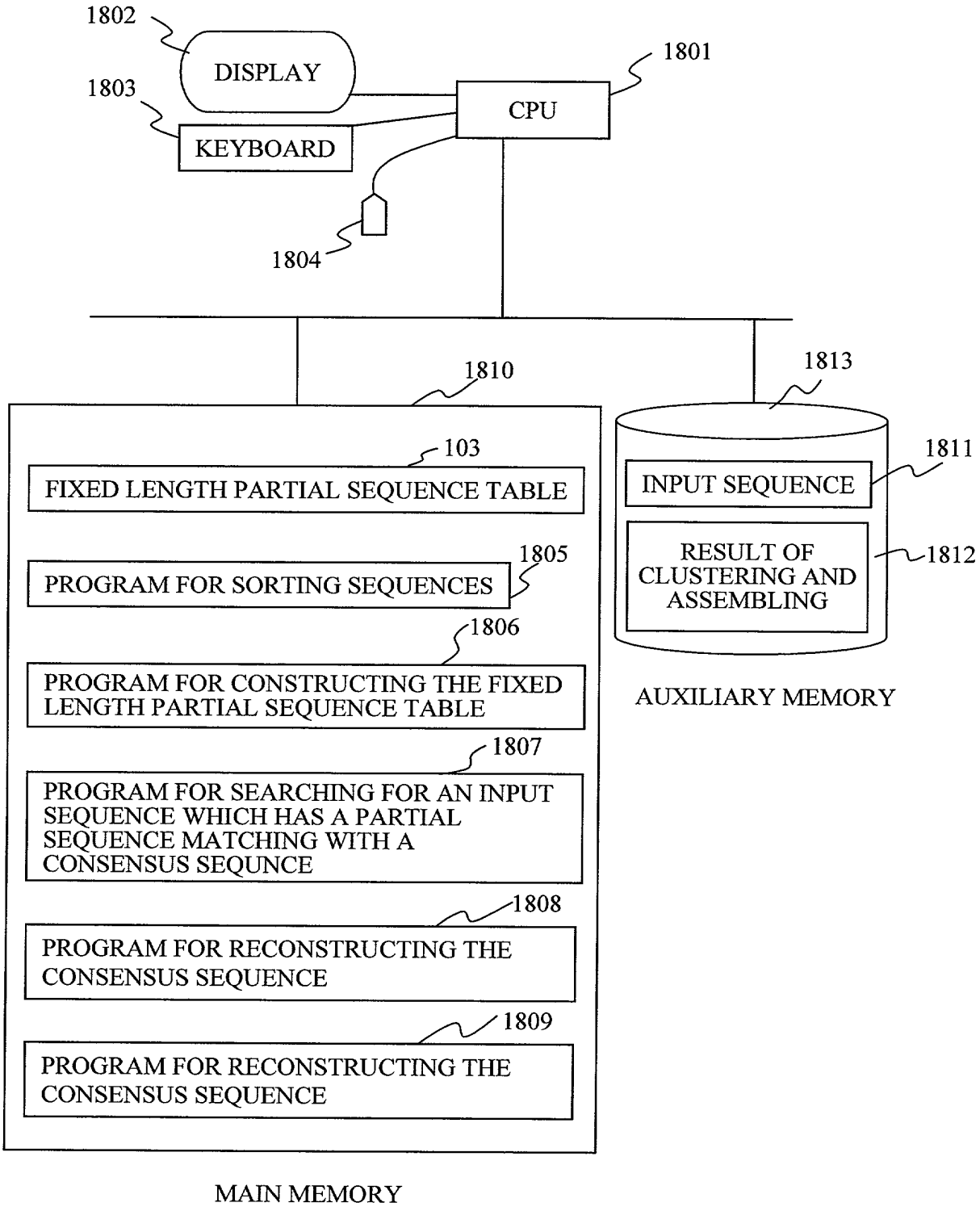


FIG.18


1802
DISPLAY
1803
KEYBOARD
1801
CPU
1804
1810
1813
103
FIXED LENGTH PARTIAL SEQUENCE TABLE
1805
PROGRAM FOR SORTING SEQUENCES
1806
PROGRAM FOR CONSTRUCTING THE FIXED LENGTH PARTIAL SEQUENCE TABLE
1807
PROGRAM FOR SEARCHING FOR AN INPUT SEQUENCE WHICH HAS A PARTIAL SEQUENCE MATCHING WITH A CONSENSUS SEQUNCE
1808
PROGRAM FOR RECONSTRUCTING THE CONSENSUS SEQUENCE
1809
PROGRAM FOR RECONSTRUCTING THE CONSENSUS SEQUENCE
MAIN MEMORY
1811
INPUT SEQUENCE
1812
RESULT OF CLUSTERING AND ASSEMBLING
AUXILIARY MEMORY